ADDIS

Zhao, Jing; van Valkenhoef, Gert; de Brock, E.O.; Hillege, Hans

university of groningen

faculty of economics and business

12007-Other

ADDIS: an automated way to do network meta-analysis

Jing Zhao
Gert van Valkenhoef
Bert de Brock
Hans Hillege

SOM is the the research institute of the Faculty of Economics & Business at the University of Groningen. SOM has six programmes:
- Economics, Econometrics and Finance
- Global Economics & Management
- Human Resource Management & Organizational Behaviour
- Innovation & Organization
- Marketing
- Operations Management & Operations Research

Research Institute SOM
Faculty of Economics & Business
University of Groningen

Visiting address:
Nettelbosje 2
9747 AE Groningen
The Netherlands

Postal address:
P.O. Box 800
9700 AV Groningen
The Netherlands

T +31 50 363 7068/3815

www.rug.nl/som

ADDIS: an automated way to do network meta-analysis

Jing Zhao
University of Groningen/UMCG
j.zhao.5@student.rug.nl

Gert van Valkenhoef
University of Groningen

Bert de Brock
University of Groningen

Hans Hillege
University of Groningen/UMCG

ADDIS: an automated way to do network meta-analysis

Jing Zhao[*1], Gert van Valkenhoef[1,2], Bert de Brock[2], Hans Hillege[1]

July 19, 2012

[1] Department of Epidemiology, University of Groningen, University Medical Center Groningen, The Netherlands

[2] Faculty of Economics and Business, University of Groningen, The Netherlands

Email: Jing Zhao - j.zhao.5@student.rug.nl; Gert van Valkenhoef - g.h.m.van.valkenhoef@rug.nl; Bert de Brock - e.o.de.brock@rug.nl; Hans Hillege - h.hillege@umcg.nl

[*] Corresponding author

1 Introduction

In evidence-based medicine, meta-analysis is an important statistical technique for combining the findings from independent clinical trials which have attempted to answer similar questions about treatment's clinical effectiveness [1]. Normally, such meta-analyses are pair-wise treatment comparisons, which only include the comparisons between two treatments, e.g. treatment A and placebo. When additional treatments are of interest (e.g. treatment B and treatment C), pair-wise treatment comparison starts showing its limitations as it only accesses the evidence from direct comparisons between two treatments and can not guarantee consistency between comparisons. Network meta-analysis is a statistical method for combining both direct and indirect evidence from multiple trials in order to obtain a single consistent quantitative synthesis [2, 3, 4]. It enables to detect the heterogeneity among different trials comparing the same treatments and inconsistency between direct and indirect evidence.

Compared to pair-wise meta-analysis, network meta-analysis is rather difficult to conduct due to the need for analyzing inconsistency, specifying the model, assessing convergence, etc. The purpose of this report is to introduce an automated way to perform network meta-analysis through ADDIS (Aggregate Data Drug Information System).

ADDIS is a decision support system developed to assist researchers to store data from clinical trials in a structured way and create meta-analyses, network meta-analyses, and benefit-risk assessments in an easy and user-friend way [5]. It is open source software and can be download at http://drugis.org/addis for free. Algorithms have been developed for ADDIS to automatically generate consistency models [6], inconsistency models [7], and node-splitting models [8]. Graphical and quantitative summaries are provided to assist the interpretation and improve the presentation of results from network meta-analysis. To illustrate how a network meta-analysis can be performed in ADDIS, we re-analyze the data from a published systematic review and network meta-analysis in antihypertensive therapies [9].

2 Background

Network meta-analysis statistically combines the results of a collection of clinical trials comparing all or some of the treatments of interest. The key assumption underlying any meta-analysis is exchangeability of the trials: the trials all measure the same underlying relative treatment effects; any observed differences are due to chance. To apply this definition to networks in which trials do not all compare the same treatments, we assume that the missing arms are missing "at random". Suppose there are three treatments, A, B, and C, and pair-wise comparisons AB and AC. Then, the assumption is that if an AB trial would have also included a C arm, it would measure the same underlying relative effect for AC as the AC trials included in the network.

The best way to assess exchangeability is to collect information about the clinical trials, and carefully consider whether they are indeed similar enough to be compared. For example, are all the trials of similar duration, or are the trials that compare A and B much shorter than those that compare B and C? If we find such a difference, this could have an important impact on whether the results of a network meta-analysis are reliable. Other potential confounding factors include dosage, study quality, patient inclusion and exclusion criteria, publication bias, etc. Thus, when defining the dataset for a network meta-analysis, we should be careful to ensure that the included clinical trials are as homogeneous as possible.

If the exchangeability assumption is violated in a pair-wise meta-analysis, this may lead to heterogeneity: a larger than expected variation in results between trials. When the amount of heterogeneity is large, it may be inappropriate to calculate an overall summary of effect size. On the other hand, there are often small differences in the population or conduct of trials that lead to a minor amount of heterogeneity. In this case, a relaxed version of the exchangeability assumption may be more appropriate: we allow for some variation between trials, and attempt to model it. Thus, rather than each trial measuring the same underlying effect (a fixed effects model), the underlying effect of each trial is drawn from a normal distribution with a common mean and variance (a random effects model). In a random effects model, the random effects variance is

an important quantity that measures the heterogeneity between trials.

In a network meta-analysis, a violation of exchangeability can also lead to differences between (rather than within) comparisons. This is called inconsistency (and sometimes incoherence): disagreement between direct and indirect evidence. Note that inconsistency can only be detected when both direct and indirect evidence are available for a comparison. Thus, if we compare AC indirectly through AB and BC there could be a systematic difference between the AB and the BC trials that leads to inconsistency, but that inconsistency can not be detected statistically in the absence of AC trials. There are several statistical models to detect inconsistency (discussed below), but it is important to note that even if such a model does not detect inconsistency, that does not mean that the network indeed consists of exchangeable trials. The most important step to ensure valid results remains the careful selection of trials for inclusion in the analysis.

2.1 Consistency analysis

In pair-wise meta-analysis, we could derive three unrelated estimates of relative effects d_{AB}, d_{AC}, and d_{BC} from the given trials. In network meta-analysis, by contrast, we estimate all three relative effects simultaneously through the consistency constraint: $d_{BC} = d_{AC} - d_{AB}$. This means that the parameter d_{BC} is estimated from both direct evidence on BC and indirect evidence on AC and AB. When doing a network meta-analysis, one is normally only interested in the results under consistency, since only consistent results provide a solid basis for decision making. However, we can not blindly apply the consistency model, because doing that will force the results to be consistent, even when the data are not. If that is the case, we could draw incorrect conclusions from the consistency model.

2.2 Inconsistency analysis

Inconsistency [3] or node-splitting models [10] can be applied to check whether the trials in the network are indeed consistent. Both of them are based on relaxing the consistency constraint. In a random effects model, small differences between comparisons could also be modeled as heterogeneity. Thus, if a model with the consistency constraint has much higher random effects variance than the individual pair-wise comparisons, that indicates a potential inconsistency problem.

The inconsistency model assesses inconsistency by adding inconsistency factors to closed loops: $d_{BC} = d_{AC} - d_{AB} + \phi$, where ϕ is an inconsistency factor, representing the discrepancy between the direct and indirect evidence. The number of independent (potential) inconsistencies determines the number of inconsistency factors. Each inconsistency factor lies on a cycle (e.g. ABC) rather than on individual pair-wise comparisons. The size of an inconsistency factor reflects the inconsistency within the cycle [3]. The inconsistency factors share a common variance, the inconsistency variance. It has been suggested that if

the inconsistency variance is greater than the random effects variance, that indicates an inconsistency problem [3]. The inconsistency model can assess all potential inconsistencies in a single model, but is difficult to interpret and may lack statistical power.

By contrast, a different node splitting model must be estimated for each comparison that involves both direct and indirect evidence. This is more time-consuming, but the results are easier to interpret. The node splitting model selects a single comparison for which the direct and indirect evidence are compared [10]. Two posterior distributions are obtained for the mean treatment effect, e.g. d_{BC}: one, d_{BC}^{Dir}, based on pair-wise analysis of the trials including B and C, and another, d_{BC}^{Ind}, based on indirect comparison through a consistency model of the trials including A and B and the trials including A and C. The difference between the two, $d_{BC}^{Dir} - d_{BC}^{Ind}$, is called inconsistency parameter. If the two trials are exchangeable, the corresponding inconsistency parameter should be zero.

2.3 Convergence diagnostics

Network meta-analysis models in ADDIS are implemented in the Bayesian framework and estimated using Markov chain Monte Carlo (MCMC) methods. This approach is recommended by the National Institute for Health and Clinical Excellence (NICE) Decision Support Unit technical support documents on evidence synthesis [11] and commonly used in the literature [3, 4, 10, 12]. MCMC methods [13] are extremely flexible in estimating statistical models, but their application requires some care. The methods work by initially giving the model's parameters some arbitrary values (the *starting point*) and then updating the parameters each iteration using some stochastic process. In this manner, the parameters (samples) generated each iteration are correlated with the samples generated the previous iteration; this is called a Markov chain. Eventually, the Markov chain will provide an accurate estimate of the statistical model, in which case we say it has *converged*. However, how many iterations are required to get an accurate estimate is not known in advance.

To determine whether or not sufficient iterations have been generated, convergence diagnostics have to be applied. The Brooks-Gelman-Rubin diagnostic [14, 15] implemented in ADDIS runs several Markov chains with different starting points and compares their results. When all the chains are similar, this indicates that the model has converged. Specifically, convergence is assessed by comparing within-chain and between-chain variance to calculate the Potential Scale Reduction Factor (PSRF) [15]. If the PSRF is large, it means that the between-chains variance can be decreased by running additional iterations. If the PSRF is close to 1, it indicates approximate convergence has been reached. An iterative PRSF plot is useful to verify that the between-chain variance is decreasing and the within-chain variation is increasing as the simulations converge [15].

3 Example dataset

Psaty et al [9] performed a network meta-analysis which combined clinical trial data from 42 studies that included 192,478 patients randomized to 7 major treatment strategies for patients with uncomplicated hypertension. The treatment strategies evaluated in their network meta-analysis are:

- Placebo
- Diuretic therapy (chlorothiazide or hydrochlorothiazide 12.5 to 25 mg/day)
- β-blockers
- Angiotensin-converting enzyme (ACE) inhibitors
- Angiotensin receptor blockers (ARB)
- Calcium channel blockers (CCB)
- α-blockers

Six clinical outcomes are assessed:

- Coronary heart disease (CHD)
- Congestive heart failure (CHF)
- Stroke
- Cardiovascular disease (CVD) events
- CVD mortality
- Total mortality

The network of all the treatments from included studies is shown in Figure 1.

4 Implementation in ADDIS

In this section, we show how to reproduce the network meta-analysis by Psaty et al [9] in ADDIS. We start with a step-by-step instruction for defining the dataset for the network meta-analysis. Then we illustrate the consistency analysis and inconsistency analysis for the outcome "Coronary Heart Disease". Next we explain how the assessment of convergence can be done in ADDIS. Finally, after concluding that there is no discernible inconsistency, we show the results on all six outcomes based on the results of consistency analysis as Forest plots relative to Diuretics.

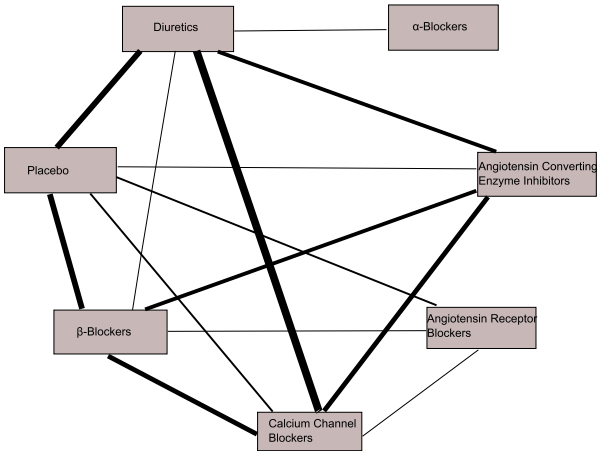


Figure 1: Network meta-analysis of first-line anti-hypertensive drug treatments. The width of the lines indicates the number of studies that include that comparison (the minimum is 1 and the maximum is 8)

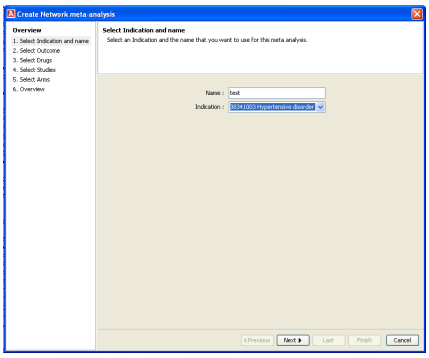
4.1 Defining the analysis dataset

ADDIS assists implementation of network meta-analysis based on the entered data (a tutorial of entering data in ADDIS is at http://drugis.org/addis1.8). The dataset for this example can be obtained by "Load example" in ADDIS. The process of defining a network meta-analysis dataset is presented as a series of screen shots in Figure 2. By clicking "New Network meta-analysis" on the main window of ADDIS, the user starts this step wise process.

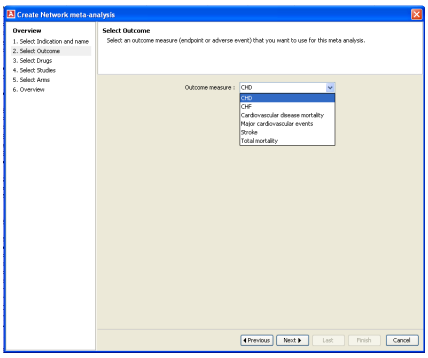
Step 1: Choose a name and an indication (Figure 2 (a)). The user gives a name to the analysis (e.g. "Psaty CHD") and chooses an indication ("38341003 Hypertensive disorder" in this example). Based on the chosen indication, the system selects and presents all the outcome measures from different studies with this indication.

Step 2: Choose an outcome (Figure 2 (b)). The user selects an outcome ("CHD" in this example) that he or she would like to analyze. Based on the chosen outcome, the system selects the studies and treatments that can be included in the analysis.

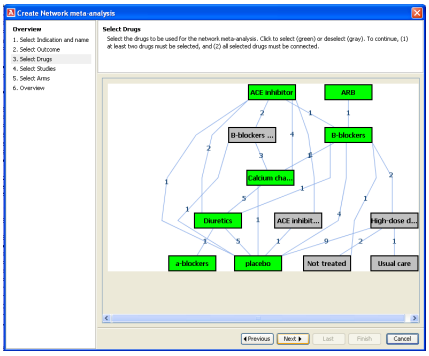
Step 3: Choose alternatives (Figure 2 (c)). Based on the available treatments, the system constructs the evidence graph, where the user can pick the ones to include in the analysis. A green block means that the treatment is included and a gray one that it is excluded. The number on the line represents the number of studies participating in each comparison. The system will not continue if the selected treatments do not form a connected evidence network. Placebo, Diuretics, ACE inhibitor, CCB, ARB, α-blockers, and β-blockers were selected in this example.



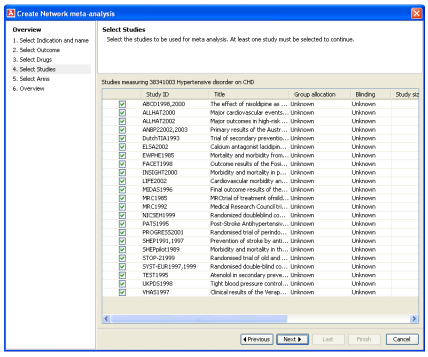
(a) Select indication and name



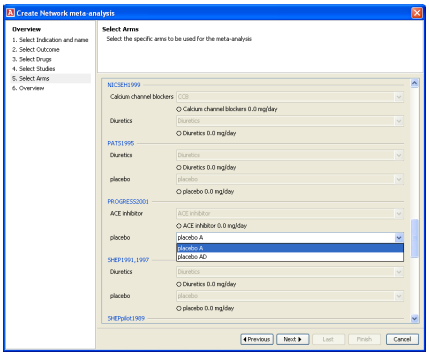
(b) Select outcome



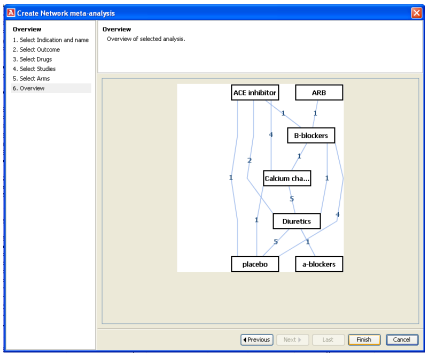
(c) Select drugs



(d) Select studies



(e) Select arms



(f) Overview

Figure 2: Step-wise instruction of defining the dataset for network meta-analysis in ADDIS

Step 4: Choose studies (Figure 2 (d)). The system lists all the studies measuring at least two of the selected treatments on the selected outcome, and the user can exclude the ones not desired based on their characteristics. Since

the database contains only the studies included in the systematic review [9], all the studies were included in this example.

Step 5: Choose treatment arms (Figure 2 (e)). When more than one matching arm is available, the user must choose the appropriate one in this step based on the arm's characteristics. The user can go back to step 4 to exclude the study if no appropriate arm is available. In this example, "placebo A" was chosen in study PROGRESS2001 as it is the placebo compared to ACE inhibitor which is included in the analysis.

Step 6: Overview (Figure 2 (f)). The final evidence network is shown in the overview, and if the network is not connected, it is not possible to finish this step. In that case, the user needs to go back to step 3 to exclude some of the treatments. By clicking "Finish", the system saves the analysis model.

The saved analysis, "Psaty CHD", can be found under "network meta-analysis" on the left side of ADDIS main screen. By selecting it, the user will get an overview of the network meta-analysis dataset he or she just created.

4.2 Consistency analysis

To run the consistency model, the user should select the section "Consistency" and then click the run button. Convergence of the model should be assessed before drawing conclusions based on the results (see Section 4.4). We extended the number of iterations once in order to achieve good convergence (by clicking "No, extend" once). Table 1 gives the odds-ratios and the 95% Credibility Interval (CrI) for all treatments relative to each other under the consistency model. For example, the odds-ratio between ARB and Placebo is 0.94 with Placebo as baseline and 1.07 with ARB as baseline, so incidence of CHD is lower for ARB (not significant). The median random effects variance is 0.15 with 95% CrI (0.03, 0.34). Table 2 gives the probability of each alternative to obtain each rank [4]. Rank 1 is the worst indicating the highest incidence of CHD, and rank 7 is the best indicating the lowest incidence of CHD. For example, according to this rank probability (Table 2), ACE inhibitor and α-blockers are better alternatives compared to the other treatments as they have much higher score on rank 7 (0.38 and 0.33 respectively), which indicates they have much lower incidence of CHD. On the contrary, Placebo is the worst with highest rank 1 probability (0.47) and lowest rank 7 probability (0.0). The rank probabilities sum to one, both within a rank over all treatments and within a treatment over all ranks. The results are also visualized in Figure 3.

4.3 Inconsistency analysis

In our example, trial selection was already taken care of by the systematic review [9], and thus we could include all relevant studies without further consideration. Thus, we go directly to statistical modeling to detect inconsistency.

Table 1: Relative effects table from the consistency model. Odds-ratio (95% Credibility Interval). ACE indicates angiotensin-converting enzyme; CCB indicates calcium channel blockers; and ARB indicates angiotensin receptor blockers.

ACE inhibitor	1.26 (0.77, 2.06)	1.19 (0.90, 1.58)	1.18 (0.95, 1.47)	1.03 (0.83, 1.26)	1.04 (0.67, 1.60)	1.34 (1.05, 1.72)
0.80 (0.48, 1.31)	ARB	0.95 (0.63, 1.58)	0.94 (0.57, 1.55)	0.82 (0.50, 1.33)	0.82 (0.44, 1.53)	1.07 (0.67, 1.69)
0.84 (0.63, 1.11)	1.05 (0.70, 1.58)	β-blockers	0.99 (0.74, 1.32)	0.86 (0.66, 1.12)	0.87 (0.55, 1.37)	1.12 (0.90, 1.40)
0.85 (0.68, 1.06)	1.07 (0.65, 1.77)	1.01 (0.76, 1.36)	CCB	0.87 (0.70, 1.09)	0.88 (0.57, 1.36)	1.14 (0.88, 1.47)
0.97 (0.79, 1.20)	1.23 (0.75, 1.99)	1.16 (0.90, 1.51)	1.15 (0.92, 1.43)	Diuretics	1.01 (0.69, 1.47)	1.31 (1.06, 1.62)
0.96 (0.63, 1.48)	1.21 (0.65, 2.25)	1.15 (0.73, 1.82)	1.14 (0.73, 1.76)	0.99 (0.68, 1.44)	α-blockers	1.30 (0.84, 1.99)
0.74 (0.58, 0.95)	0.94 (0.59, 1.49)	0.89 (0.72, 1.11)	0.88 (0.68, 1.14)	0.76 (0.62, 0.94)	0.77 (0.50, 1.19)	Placebo

Table 2: Probability for each alternative to be at each rank given the analysis model and the data. Rank 1 means highest incidence of Coronary Heart Disease and rank 7 the lowest. ACE indicates angiotensin-converting enzyme; CCB indicates calcium channel blockers; and ARB indicates angiotensin receptor blockers.

Drug	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6	Rank 7
ACE inhibitor	0.00	0.01	0.03	0.07	0.21	0.30	0.38
ARB	0.35	0.19	0.15	0.10	0.06	0.06	0.09
β-Blockers	0.05	0.23	0.37	0.19	0.10	0.05	0.02
CCB	0.07	0.17	0.21	0.35	0.15	0.05	0.01
Diuretics	0.00	0.01	0.03	0.10	0.30	0.39	0.17
α-Blockers	0.06	0.07	0.08	0.14	0.17	0.15	0.33
Placebo	0.47	0.33	0.14	0.04	0.01	0.00	0.00

4.3.1 Node split model

To run the node split model, the user should select the section "Node Split" and then click "Run all node-split models" or click the run button for each node split model separately. Convergence of each model should be assessed before drawing conclusions based on the results (see Section 4.4). Table 3 lists the relative effect from direct and indirect evidence and the significance level of the evidence. The p-values show no significant difference between the direct effect and indirect effect. We present the density plots of the best comparison (β-blockers vs. Placebo with p=0.95) and the worst comparison (ACE inhibitor vs. β-blockers with p=0.09) in Figure 4a and 4b, respectively. We can see that in the best comparison, the relative effect plots from direct evidence, indirect evidence, and consistency model resemble each other, while the ones in the worst

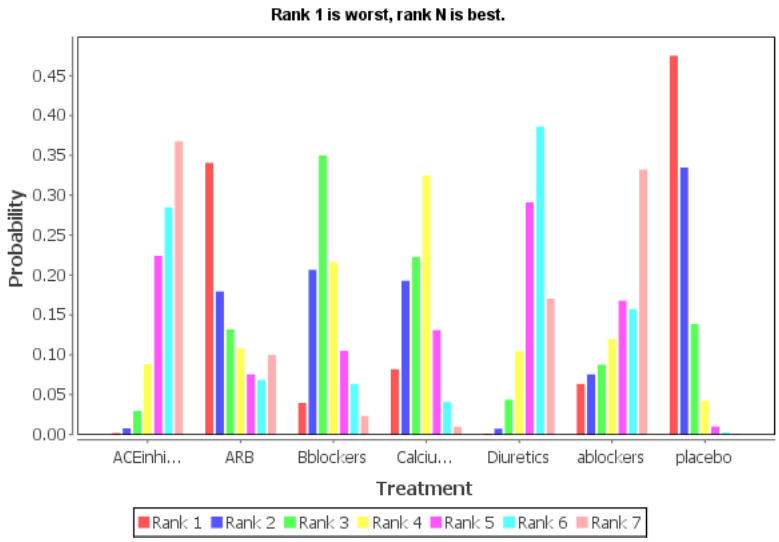
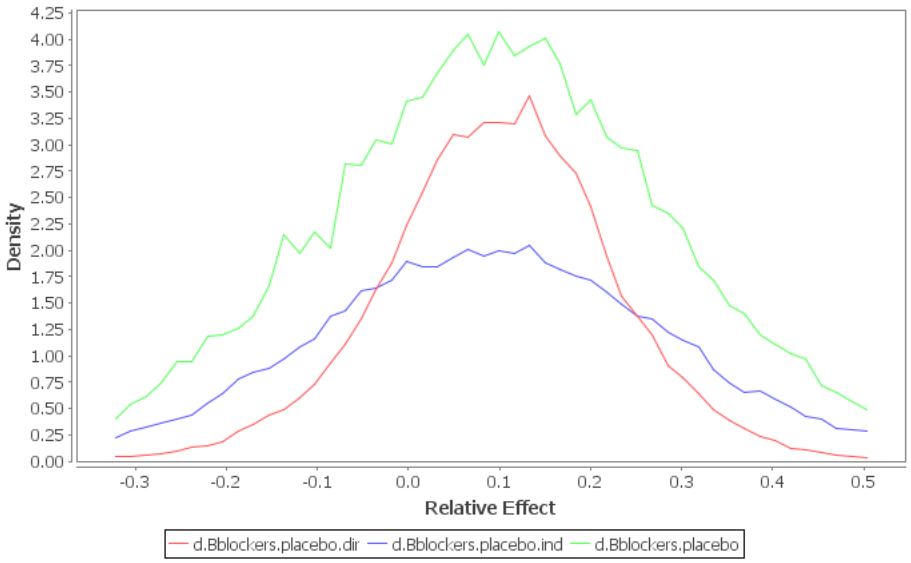


Figure 3: Visualized rank probabilities. ACE indicates angiotensin-converting enzyme and ARB indicates angiotensin receptor blockers.

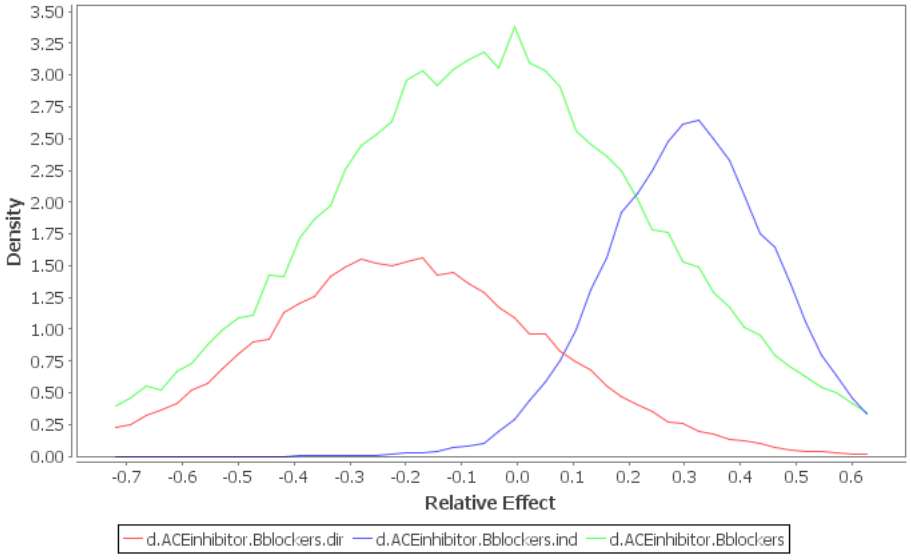
comparison are more divergent. In general, the results from the node split model show no significant inconsistency is present in this dataset.

Table 3: Node splitting model. Log odds-ratio (95% credibility interval). ACE indicates angiotensin-converting enzyme; CCB indicates calcium channel blockers; and ARB indicates angiotensin receptor blockers.

Name	Direct effect	Indirect effect	Overall	P
d.ACE inhibitor.Diurectics	0.06(-0.17,0.31)	-0.16(-0.43,0.17)	0.02(-0.19,0.24)	0.23
d.ACE inhibitor.Placebo	0.09(-0.46,0.62)	0.36(0.08,0.64)	0.29(0.04,0.54)	0.38
d.β-blockers.ACE inhibitor	0.20(-0.32,0.74)	-0.31(-0.64,-0.01)	-0.18(-0.45,0.10)	0.09
d.β-blockers.Diuretics	-0.47(-0.97,-0.01)	-0.05(-0.34,0.22)	-0.16(-0.41,0.11)	0.11
d.β-blockers.Placebo	0.10(-0.17,0.35)	0.09(-0.33,0.51)	0.11(-0.11,0.33)	0.95
d.CCB.ACE inhibitor	-0.22(-0.56,0.02)	0.02(-0.37,0.43)	-0.15(-0.40,0.04)	0.27
d.CCB.β-blockers	-0.03(-0.80,0.74)	0.03(-0.33,0.32)	0.03(-0.30,0.29)	0.89
d.CCB.Diuretics	-0.06(-0.36,0.21)	-0.20(-0.61,0.13)	-0.13(-0.38,0.07)	0.48
d.CCB.Placebo	0.31(-0.26,0.90)	0.09(-0.26,0.36)	0.14(-0.15,0.37)	0.47
d.Diuretics.Placebo	0.30(0.04,0.56)	0.25(-0.08,0.59)	0.27(0.05,0.47)	0.80



(a) β-blockers vs. Placebo



(b) ACE inhibitor vs. β-blockers

Figure 4: Density plot of direct and indirect evidence . The red line represents the evidence from direct comparisons, the blue line represents the evidence from indirect comparisons, and the green line represents the evidence from the consistency analysis. ACE indicates angiotensin-converting enzyme.

4.3.2 Inconsistency model

To run the inconsistency model, the user should select the section "Inconsistency" and then click the run button. Convergence of the model should be assessed before drawing conclusions based on the results (see Section 4.4). We extended the number of iterations twice in order to achieve good convergence (by clicking "No, extend" twice). The inconsistency factors (Table 4) were calculated as median and 95% credibility interval. From the results we can see that all the intervals include 0 which suggests no significant inconsistency. Even in the most extreme inconsistency factor, 0.08 (-0.16, 0.63), the median value is within one standard deviation from zero. The inconsistency variance is not much different from the random effects variance (Table 5). Therefore, the results show no relevant inconsistency, which indicates that the studies included in these comparisons did not lead to inconsistencies.

In addition, the inconsistency model also provides a relative effects table which can be compared to the one from the consistency model (Table 1). In this example, the relative effects from inconsistency model are comparable to the ones from consistency table, indicating no relevant inconsistency. Note that the consistency assumption is critical for decision making and so the results of an inconsistency model should *never* be used to draw conclusions about the relative effects of the included treatments.

Table 4: Inconsistency factors. CrI indicates credibility interval; ACE indicates angiotensin-converting enzyme; CCB indicates calcium channel blockers; and ARB indicates angiotensin receptor blockers.

Cycle	Median (95% CrI)
ACE inhibitor, β-blockers, CCB	-0.07(-0.54,0.17)
ACE inhibitor, CCB, β-blockers, Diuretics	-0.06(-0.56,0.21)
ACE inhibitor, CCB, β-blockers, Placebo, Diuretics	0.02(-0.37,0.42)
β-blockers, CCB, Diuretics	0.08(-0.16,0.63)
β-blockers, CCB, Diuretics, Placebo	0.01(-0.29,0.38)
CCB, Diuretics, Placebo	-0.03(-0.47,0.27)

Table 5: Variance calculation. CrI indicates credibility interval.

Parameter	Median (95% CrI)
Random Effects Variance	0.15(0.04,0.34)
Inconsistency Variance	0.17(0.01,0.66)

4.4 Convergence diagnostics

When ADDIS asks to assess convergence (Figure 5), the user should click the button "show convergence". Then, ADDIS will show the convergence assessment dialog (Figure 6). If convergence is reached, the user could click "Yes,

finish" to finish the MCMC simulation. Otherwise, the user could click "No, extend" to let the model run for more iterations until convergence is adequate. The PRSF in the convergence parameter table is just a snapshot result at the end of the simulation. To check whether the convergence is reached stably or only temporarily, we need to look at the convergence plots. By double clicking a parameter in the convergence table, the user can obtain the convergence plots. For example, Figure 6 shows the convergence table of the consistency model and Figure 7 shows the convergence plot of the parameter of β-blockers vs. ACE inhibitors. From the plot we can see that the PSRF is quite far away from one in the beginning before approximate convergence has been reached and close to one after 5,000 iterations, and the convergence is very stable after 50,000 iterations.

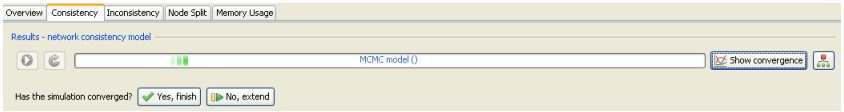


Figure 5: The progress indication after the simulation phase has been completed. The user is prompted to assess convergence.

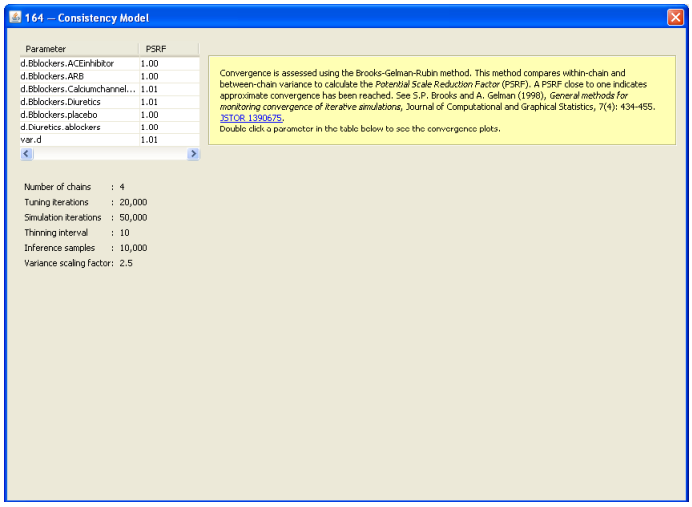


Figure 6: Convergence parameter table from the consistency model.

4.5 Overall results

As there is no relevant inconsistency in the evidence (as determined by Psaty et al [9] for the remaining outcomes), we can use the consistency model to draw conclusion about the relative effects of the included treatments. The relative

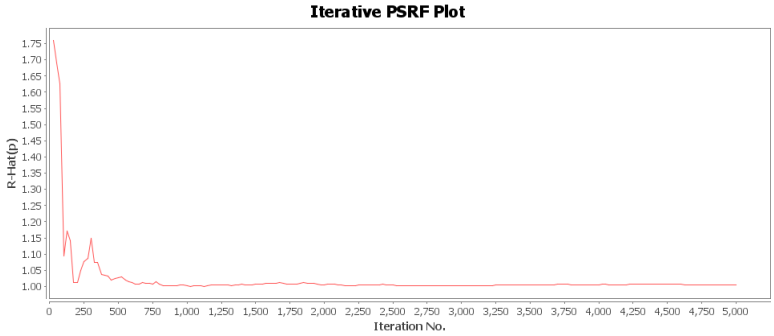


Figure 7: Convergence plot of chain β-blockers and Angiotensin-Converting Enzyme inhibitors. PSRF indicates potential scale reduction factor. The iteration number is thinned by factor 10, so for example iteration 5,000 in the figure actually corresponds with iteration 50,000.

effects (odds-ratio) between Diuretics and other anti-hypertensive therapies are plotted in Figure 8 based on a consistency analysis in ADDIS

5 Discussion

In this report, we introduced ADDIS as an automated way to perform network meta-analysis. ADDIS is an integrated software application providing decision support based on clinical trial results. Besides creating network meta-analysis, ADDIS also enables the construction of pair-wise meta-analysis and benefit-risk decision models.

For network meta-analysis, ADDIS supports both continuous and dichotomous outcomes, and provides two ways to assess inconsistency. The user interface facilitates convergence checking and interpretation of the results, with figures and tables tailored to each model.

5.1 Analysis of the example dataset

To illustrate the application of ADDIS for network meta-analysis, the aggregated data from a systematic review were used. The inconsistency analysis shows no significant inconsistency between the direct evidence and indirect evidence, therefore we assume that the consistency model is valid. Otherwise, there would be a need to investigate the source for inconsistency from the involved studies and exclude them and run the model until there is no significant inconsistency.

Psaty et al calculated the logarithm of the risk-ratio (RR), while ADDIS calculates the logarithm of the odds-ratio (OR). The OR and RR are similar when the event is rare, but differ substantially when the event is common. However, in network meta-analysis, the log OR is more suitable than the log RR because the former has better mathematical properties and often reflects the

underlying mechanisms more effectively. It does not matter whether we analyze the event or non-event with OR, but with RR it requires care in choosing event or non-event. This is because the OR of an event is reciprocal with the OR of the corresponding non-event, which makes the calculation much easier.

The relative effects based on consistency model are consistent between the results from ADDIS and the ones from the systematic review. There are some small differences due to different ways to deal with some treatment arms. The differences mainly focus on Diuretics vs. β-Blockers (Figure 8b) and Diuretics vs. ARB (Figure 8e) on outcome CHF. Although neither of them reaches a significant level, the results show the opposite favorable alternative. The difference on Diuretics vs β-Blockers is due to the way the treatment arm "β-blocker or diuretics" was analyzed. There are five studies including "β-blocker or diuretics" as a single treatment group in the systematic review. This treatment arm was assessed as two separate treatments by weighing 68% to β-blocker and 32% to diuretics because this allocation was reported in one study. However, how the other four studies treat it is unknown. Therefore, this arm was excluded in our network meta-analysis. This could also influence the results between Diuretics and ARB because there is no direct evidence comparing them. ARB was only compared directly with Placebo, β-Blockers and CCB. Therefore the results in Figure 8e are estimated through indirect evidence, such as, ARB is compared with β-Blockers and β-Blockers is compared with Diuretics. Since the comparison between β-Blockers and Diuretics is influenced by the mixed arm "β-blocker or diuretics", the comparison between ARB and Diuretics is also influenced.

5.2 Future work

In the future, ADDIS will enable the analysis of more complicated data, such as survival data (time-to-event data), and will enable the generation of Forest plots for relative effects on pair-wise comparisons. It will be possible to stratify treatments by dose. For example, currently in ADDIS, treatment A is defined as drug X, but in the future, treatment A can be defined as low-dose drug X and treatment B can be defined as high-dose drug X.

Future work will also include additional methods to assess convergence and run length, new tools to make it easier to identify the studies that are causing inconsistency (e.g. independent mean effects model), tools for study selection and treatment definition, and meta-regression to model the effect of covariates (e.g. to correct for trial duration or dosage).

References

[1] M. Egger, G. D. Smith, and A. N. Phillips. Meta-analysis: principles and procedures. *BMJ*, 315:1533, 1997.

[2] T. Lumley. Network meta-analysis for indirect treatment comparisons. *Statistics in Medicine*, 21(16):2313–2324, 2002.

[3] G. Lu and A. E. Ades. Assessing evidence inconsistency in mixed treatment comparisons. *Journal of American Statistical Association*, 101(474):447–459, 2006.

[4] G. Salanti, J. P. T. Higgins, A. E. Ades, and J. P. A. Ioannidis. Evaluation of networks of randomized trials. *Statistical Methods in Medical Research*, 17(3):279–301, 2008.

[5] G. van Valkenhoef, T. Tervonen, T. Zwinkels, B. de Brock, and H. Hillege. Addis: a decision support system for evidence-based medicine. *Decision Support Systems*, In press, 2012.

[6] G. van Valkenhoef, G. Lu, B. de Brock, H. Hillege, A. E. Ades, and N. J. Welton. Automating network meta-analysis. *Submitted manuscript*, 2012.

[7] G. van Valkenhoef, T. Tervonen, B. de Brock, and H. Hillege. Algorithmic parameterization of mixed treatment comparisons. *Statistics and Computing*, in press, 2011.

[8] G. van Valkenhoef, S. Dias, N. S. Weiss, and A. E. Ades. Automated generation of mixed treatment comparison node-splitting models. *draft manuscript*, 2012.

[9] B. M. Psaty, T. Lumley, C. D. Furberg, G. Schellenbaum, M. Pahor, M. H. Alderman, and N. S. Weiss. Health outcomes associated with various antihypertensive therapies used as first-line agents: a network meta-analysis. *JAMA*, 289(19):2534–2544, 2003.

[10] S. Dias, N. J. Welton, D. M. Caldwell, and A. E. Ades. Checking consistency in mixed treatment comparison meta-analysis. *Statistics in Medicine*, 29:932–944, 2010.

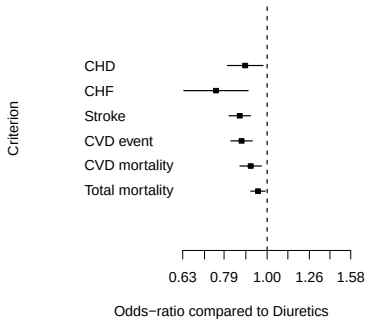
[11] S. Dias, N. J. Welton, A. J. Sutton, and A. E. Ades. Technical support documents 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomized controlled trials. Technical report, 2011.

[12] G. Lu and A. E. Ades. Combination of direct and indirect evidence in mixed treatment comparison. *Statistics in Medicine*, 23(20):3105–3124, 2004.

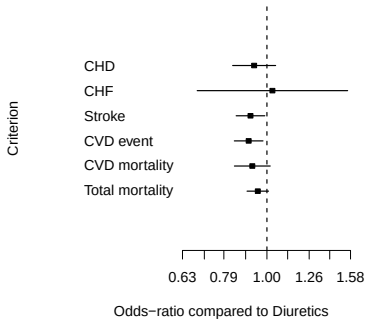
[13] C. P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer, 2nd edition, 2004.

[14] A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences. *StatSci*, 7(4):457–572, 1992.

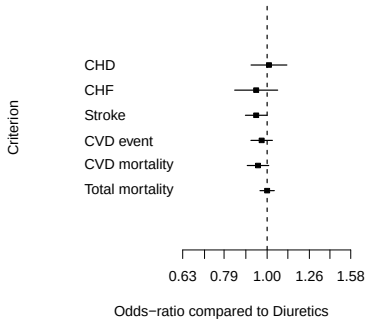
[15] S.P. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *Computational and Graphical Statistics*, 7(4):434–455, 1998.



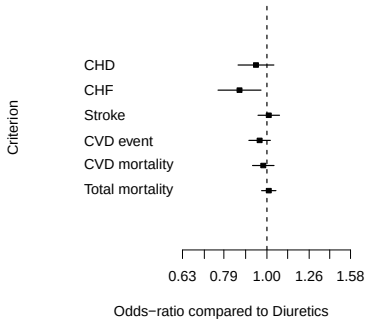
(a) Diuretics vs Placebo



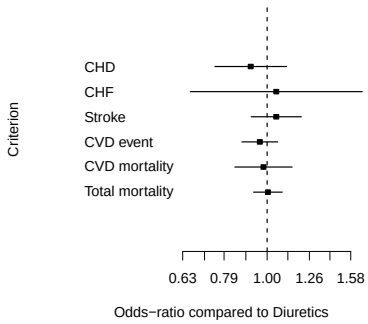
(b) Diuretics vs β-Blockers



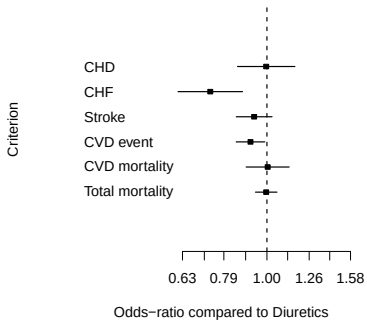
(c) Diuretics vs ACE inhibitor



(d) Diuretics vs CCB



(e) Diuretics vs ARB



(f) Diuretics vs α-Blockers

Figure 8: Relative effects between Diuretics and other treatments. Odds-ratio below one favors Diuretics. CHD indicates coronary heart disease; CHF indicates congestive heart failure; CVD indicates cardiovascular disease; ACE indicates angiotensin-converting enzyme; CCB indicates calcium channel blockers; and ARB indicates angiotensin receptor blockers.

List of research reports

12001-HRM&OB: Veltrop, D.B., C.L.M. Hermes, T.J.B.M. Postma and J. de Haan, A Tale of Two Factions: Exploring the Relationship between Factional Faultlines and Conflict Management in Pension Fund Boards

12002-EEF: Angelini, V. and J.O. Mierau, Social and Economic Aspects of Childhood Health: Evidence from Western-Europe

12003-Other: Valkenhoef, G.H.M. van, T. Tervonen, E.O. de Brock and H. Hillege, Clinical trials information in drug development and regulation: existing systems and standards

12004-EEF: Toolsema, L.A. and M.A. Allers, Welfare financing: Grant allocation and efficiency

12005-EEF: Boonman, T.M., J.P.A.M. Jacobs and G.H. Kuper, The Global Financial Crisis and currency crises in Latin America

12006-EEF: Kuper, G.H. and E. Sterken, Participation and Performance at the London 2012 Olympics

12007-Other: Zhao, J., G.H.M. van Valkenhoef, E.O. de Brock and H. Hillege, ADDIS: an automated way to do network meta-analysis

www.rug.nl/feb